

Extrakcia báz rýchlosti z toku červených krviniek ako nástroj na porovnanie modelov

autor a,b,c,d

1 Popis problému

Sú dané kanály mikrofluidného zariadenia A a B (sem by som dal obrázky). Existujú namerané toky červených krviniek - pre rôzne počty krviniek aj rôzne počiatočné podmienky. Keďže pri každom nastavení experimentu sú dráhy rozdielne - v chápaní súradníc (t, x, y, z) , nie je možné použiť priame porovnanie. Je potrebné nájsť invarianty, ktoré by bolo možné porovnať. Cieľom experimentu je nájsť bázy pre každé meranie a overiť mieru zhody pre kanál A a B.

Očakávanie : v rámci spoločného kanála sa očakáva silnejšia korelácia medzi bazami, predpokladá sa malá závislosť na počiatočných podmienkach aj počtu krviniek. Dominantný vplyv bude mať tvar kanála.

1.1 Návrh experimentov

K dispozícií bolo niekoľko dát zo simulácií tokov. Vstupom do nami navrhovaného experimentu boli pozície a rýchlosti tokov krviniek počas behu simulácie.

Prehľad je znázornený v tabuľke 1.

Tabuľka 1: Vstupné dáta

experiment	kanál	seed	počet krviniek
0	A	a	20
1	A	a	50
2	A	a	100
3	A	b	50
4	A	c	50
5	B	a	50
6	B	b	50

V prvom priblížení bude model lineárnou kombináciou báz B_i ,

$$V_j(r) = \sum_{i=1}^K w_i(r) V_{B_i} \quad (1)$$

kde

K je počet bazových funkcií,

B_i sú bázy a V_{B_i} predstavuje rýchlostnú zložku bázi,

w_i je váha i -tej bázy pre polohu krvinky r ,

V_j je predikovaná rýchlosť j -tej krvinky v polohe r ,

1.2 Učenie báz

Zvolený mechanizmus učenia báz je veľmi podobný Kohonenovým sietiam [1], kde sa váhy neurónov upravujú tak aby čo najlepšie pokrili vstupný priestor, k nim prislúchajú asociačné váhy - požadovaná hodnota výstupu. V uvedenej literatúre bola Kohonenová mapa použitá na riešenie úlohy inverznej kinematiky ramena, čo je úloha podobná s problémom nášho experimentu - hľadá sa asociácia medzi polohou a rýchlosťou v danom bode.

Pre každý experiment sa učenie spustilo nezávisle. Počiatočné hodnoty báz boli vybrané z náhodne vybraných polôh a rýchlosti krviniek. Učenie báz prebieha predkladaním polohy krvinky R_C a jej rýchlosti V_C . Bába má dve zložky polohová R_{B_i} , a rýchlostná V_{B_i} . Obe sa upravujú identickými vzťahmi (zodpovedajú metóde stochastic gradient descent [2])

$$R_{B_i} = (1 - \eta\alpha_i)R_{B_i} + \eta\alpha_i R_C \quad (2)$$

$$V_{B_i} = (1 - \eta\alpha_i)V_{B_i} + \eta\alpha_i V_C \quad (3)$$

parameter $\eta \in (0, 1)$ predstavuje rýchlosť adaptácie báz a α_i je miera podobnosti polôh R_{B_i} a R_C . Spočíta sa pomocou podobnostnej funkcie¹ podľa nasledujúcich vzťahov.

$$\beta_i = \frac{k}{k + \|R_{B_i} - R_{C_i}\|^2} \quad (4)$$

$$\alpha_i = \frac{\beta_i}{\sum_{j=1}^K \beta_j} \quad (5)$$

kde parameter $k > 0$ predstavuje strmlosť podobnostnej funkcie. Zo vzťahu je zrejmé, že $\alpha_i \in (0, 1)$, pre blízke polohy je sa hodnota blíži 1, pre vzdialené bunky sa blíži 0. Druhá rovnica predstavuje normovanie hodnôt α_i tak aby ich súčet bol 1. Zvolený počet báz bol $K = 10000$, parameter $k = 0.01$ a rýchlosť učenia $\eta = 0.1$. Zvolený počet iterácií bol 500000.

1.3 Rekonštrukcia polohy a rýchlosti bunky

Po natrénovaní báz je možné otestovať predikčné schopnosti modelu. Vstupom je ľubovoľná poloha R_T , cieľom je predpovedať aká bude rýchlosť V_T v tomto bode.

Najskôr sa spočítajú hodnoty α_i podľa vzťahov 4 a 5, ktoré korešpondujú s váhami $w_i \equiv \alpha_i$. Pomocou vzťahu 1 sa spočíta rýchlosť v danom bode.

Numerickou integráciou je možné získať trajektóriu testovanej bunky $R_T(n+1) = R_T(n) + V_C(n)dt$.

2 Výsledky

Po natrénovaní modelu bolo možné stanoviť rýchlosti v celom priereze kanála - aj v miestach kde žiadna bunka z tréningovej množiny nešla. Vďaka tomu bolo možné porovnávať experimenty medzi sebou.

Pre veľký objem dát nie vhodné odčítať všetky body, preto sa metódou Monte Carlo vybralo 1000 náhodných pozícií R_T . Výsledná vzdialenosť báz je počítaná normovanou euklidovou metrikou. Najskôr sa určia rýchlostné zložky báz dvoch natrénovaných modelov X, Y pre polohu R_T . $V_{xT} = X(R_T)$ a $V_{yT} = Y(R_T)$, kde R_T predstavuje náhodnú testovaciu polohu. Rozdiel polôh je spočítaný ako $d_T = \|V_{xT} - V_{yT}\|$. Takto vybraných 1000 pozícií je spriemerovaných. Porovnával sa každý model s každým, výsledky rozdielu sú v tabuľke 2.

Tabuľka 2: Veľkosť chyby medzi jednotlivými modelmi

	model 0	model 1	model 2	model 3	model 4	model 5	model 6
model 0	0	0.024045	0.026852	0.029994	0.028294	0.069815	0.065913
model 1	0.024174	0	0.011605	0.025831	0.023463	0.062432	0.064022
model 2	0.026618	0.011235	0	0.027454	0.026459	0.063747	0.061519
model 3	0.028648	0.025418	0.027196	0	0.022168	0.064810	0.066823
model 4	0.026980	0.026299	0.026794	0.022792	0	0.063020	0.062761
model 5	0.066236	0.067181	0.063129	0.067300	0.064338	0	0.020975
model 6	0.065729	0.066606	0.062175	0.063798	0.064817	0.022061	0

Experimenty kde sa očakáva menšia chyba (rovnaké kanály) majú spoločnú farbu - červenú pre kanál A a modrú pre kanál B. Z hodnôt je zrejmé, že modely sa líšia s menšou chybou pre rovnaký kanál. Linearita a jednoduchosť modelu však spôsobuje aj rozdiely v spoločnom kanály - príčinou sú ako aj rôzne počiatkové podmienky experimentu, ale aj fakt, že model neuvažuje kolízie buniek a nelinearitu vzťahov.

V ďalšom smerovaní sa treba zamerať na nelineárne modelovanie, napr. pomocou techník deep learning.

Ďalej sme vykonali test kde sa dáta rozdelil na dve časti - tréningovú a testovaciu (každý experiment nezávisle). Na testovanie sa zvolilo 25% dát. Merala sa odchýlka požadovanej rýchlosti od predikovanej rýchlosti modelom. Výsledky sú uvedené v tabuľke 3.

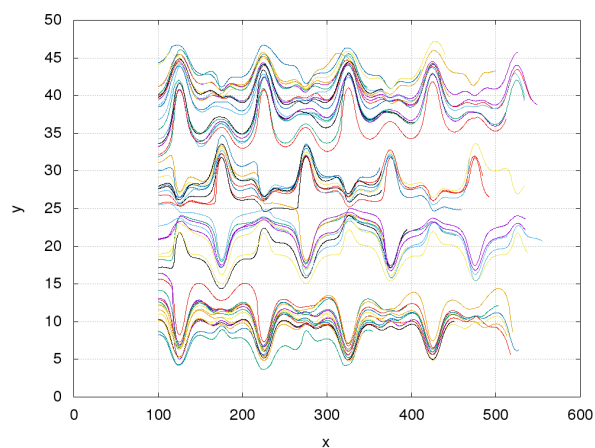
2.1 Predikcia tokov pre kanál A

Model je možné použiť na predikciu výsledku experimentu. Zvolili sme experiment 5 na otestovanie tejto hypotézy, model sme natrénovali pomocou experimentou 0 až 4, všetky pre kanál A, ale rôzne počty krviniek aj seed. Testovalo sa s iným seed (c). Požadované dráhy sú na obrázku 1 a dráhy získané z modelu sú na obrázku 2. Druhý test vznikol tréningom len na dátach s počtom krviniek 50 - experimenty 1 a 3. Je vidno že model dokáže predpovedať trajektórie. Pre jeho jednoduchosť, však dochádza aj k chybám - napr. krvinka prechádza cez pevnú prekážku, alebo nevie zachytiť zmenu prúdnice krvinky (zvíraznené oblasti obrázkov 2 a 3).

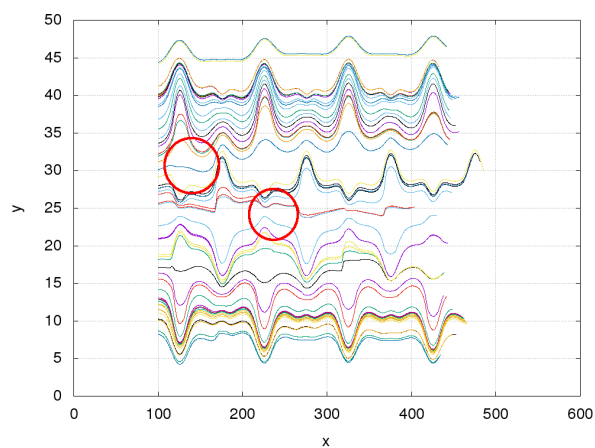
¹môže byť použitá aj iná, napr. Gaussova - časová zložitosť uvedenej je však menšia

Tabuľka 3: chyba predikcie modelov

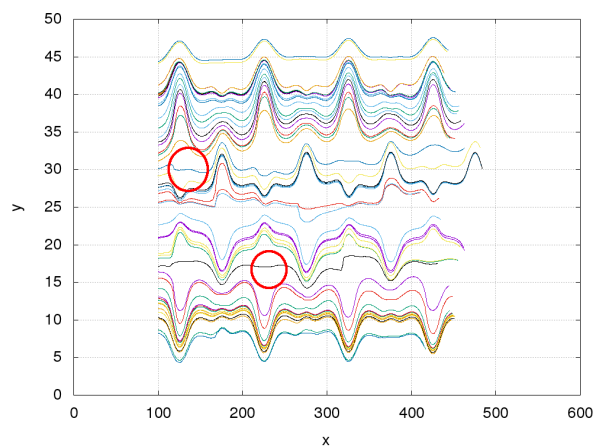
experiment	absolútna chyba	absolútna veľkosť vektora	MRE [%]
0	0.029266	0.253014	11.047184
1	0.037301	0.253148	14.865701
2	0.021849	0.131263	16.436213
3	0.038164	0.259319	14.196479
4	0.040760	0.255995	14.843164
5	0.034798	0.256949	13.244931
6	0.038682	0.246271	15.349910



Obr. 1: požadované trajektórie



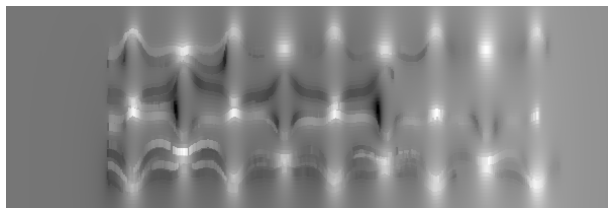
Obr. 2: získané trajektórie trénované na všetkých dátach



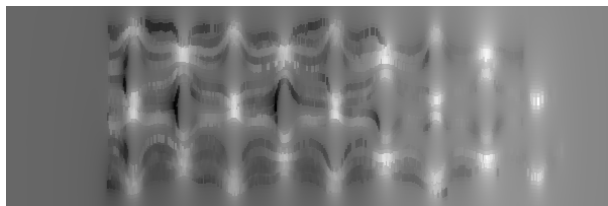
Obr. 3: získané trajektórie trénované len na dátach experimentu 1 a 3

2.2 Aproximácia tokov pre kanál A

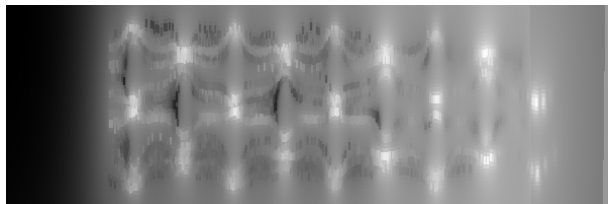
Nasledujúce obrázky predstavujú rýchlosti buniek $\|V_C\|$ v celom priereze kanála A tak ako boli získané z trénovania experimentov 0 až 4.



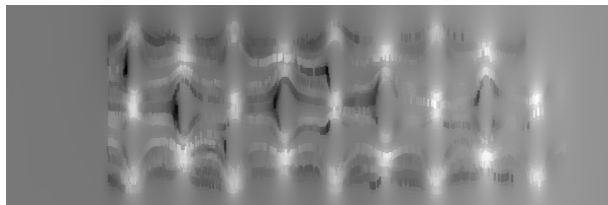
Obr. 4: experiment 0



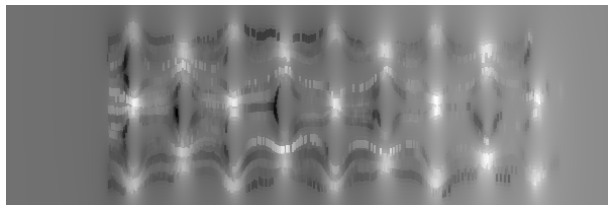
Obr. 5: experiment 1



Obr. 6: experiment 2



Obr. 7: experiment 3



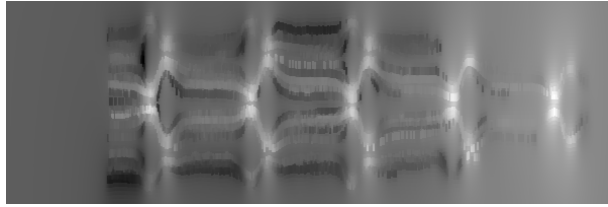
Obr. 8: experiment 4

2.3 Aproximácia tokov pre kanál B

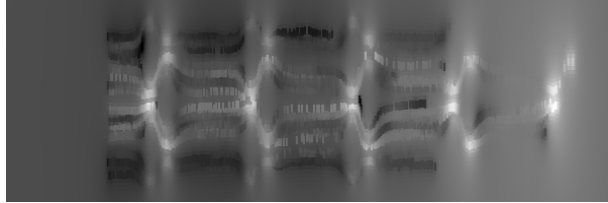
Nasledujúce obrázky predstavujú rýchlosti buniek $\|V_C\|$ v celom priereze kanála B tak ako boli získané z trénovania experimentov 5 až 6. Model umožňuje získať rýchlosť v ľubovoľnom bode.

Literatúra

- [1] R. Rojas: Neural Networks, Springer-Verlag, Berlin, 1996, Kohonen Networks



Obr. 9: experiment 5



Obr. 10: experiment 6

[2] Sebastian Ruder, An overview of gradient descent optimization algorithms